An Automatic Finite-Sample Robustness Metric: Can Dropping a Little Data Make a Big Difference?

Ryan Giordano (rgiordan@mit.edu)
December 2021

Dropping data: Motivation

Suppose you're a data analyst, and you've

- Gathered some exchangeable data,
- Cleaned up / removed outliers,
- Checked for correct specification, and
- Drawn a conclusion from your statistical analysis (e.g., based the sign / significance of some estimated parameter).

Would you be concerned if you could **reverse your conclusion** by removing a **small proportion** (say, 0.1%) of your data?

Dropping data: Mexico Microcredit

Consider Angelucci et al. [2015], a randomized controlled trial study of the efficacy of microcredit in Mexico based on 16,560 data points. The variable "Beta" estimates the effect of microcredit in US dollars.

	Left out points	Beta	(SE)
Original	0	-4.55	(5.88)

Dropping data: Mexico Microcredit

Consider Angelucci et al. [2015], a randomized controlled trial study of the efficacy of microcredit in Mexico based on 16,560 data points. The variable "Beta" estimates the effect of microcredit in US dollars.

	Left out points	Beta	(SE)
Original	0	-4.55	(5.88)
Change sign	1	0.4	(3.19)

Dropping data: Mexico Microcredit

Consider Angelucci et al. [2015], a randomized controlled trial study of the efficacy of microcredit in Mexico based on 16,560 data points. The variable "Beta" estimates the effect of microcredit in US dollars.

	Left out points	Beta	(SE)
Original	0	-4.55	(5.88)
Change sign	1	0.4	(3.19)
Change significance	14	-10.96	(5.57)

Dropping data: Mexico Microcredit

Consider Angelucci et al. [2015], a randomized controlled trial study of the efficacy of microcredit in Mexico based on 16,560 data points. The variable "Beta" estimates the effect of microcredit in US dollars.

	Left out points	Beta	(SE)
Original	0	-4.55	(5.88)
Change sign	1	0.4	(3.19)
Change significance	14	-10.96	(5.57)
Change both	15	7.03	(2.55)

Dropping data: Mexico Microcredit

Consider Angelucci et al. [2015], a randomized controlled trial study of the efficacy of microcredit in Mexico based on 16,560 data points. The variable "Beta" estimates the effect of microcredit in US dollars.

	Left out points	Beta	(SE)
Original	0	-4.55	(5.88)
Change sign	1	0.4	(3.19)
Change significance	14	-10.96	(5.57)
Change both	15	7.03	(2.55)

By removing very few data points ($15/16560 \approx 0.1\%$), we can reverse the qualitative conclusions of the original study!

Dropping data: Mexico Microcredit

Consider Angelucci et al. [2015], a randomized controlled trial study of the efficacy of microcredit in Mexico based on 16,560 data points. The variable "Beta" estimates the effect of microcredit in US dollars.

	Left out points	Beta	(SE)
Original	0	-4.55	(5.88)
Change sign	1	0.4	(3.19)
Change significance	14	-10.96	(5.57)
Change both	15	7.03	(2.55)

By removing very few data points ($15/16560 \approx 0.1\%$), we can reverse the qualitative conclusions of the original study!

Do you care? Not always. But, in some cases, surely yes!
Especially when the policy population is different than the sampled population, possibly in difficult-to-formalize ways.

Can Dropping a Little Data Make a Big Difference?

How do we find influential datapoints?

In the MX microcredit study, $\binom{16560}{15} \approx 1.4 \cdot 10^{51}$ for $\alpha = 0.0009$.

An approximation is needed!

Can Dropping a Little Data Make a Big Difference?

How do we find influential datapoints?

In the MX microcredit study, $\binom{16560}{15} \approx 1.4 \cdot 10^{51}$ for $\alpha = 0.0009$.

An approximation is needed!

We form an approximation based on the **empirical influence function**. The approximation works for all **Z-estimators** with smooth estimating equations (MLE, OLS, IV, GMM, VB, MAP, &c), and can be **computed automatically** with modern automatic differentiation.

How do we find influential datapoints?

In the MX microcredit study, $\binom{16560}{15} \approx 1.4 \cdot 10^{51}$ for $\alpha = 0.0009$.

An approximation is needed!

We form an approximation based on the **empirical influence function**. The approximation works for all **Z-estimators** with smooth estimating equations (MLE, OLS, IV, GMM, VB, MAP, &c), and can be **computed automatically** with modern automatic differentiation.

We provide finite-sample, non-stochastic accuracy guarantees. But there is no need to rely on our theory. A single re-fit provides an **exact lower bound** to data-dropping sensitivity.

Can Dropping a Little Data Make a Big Difference?

We used our R package[1] examine a number of published analyses:

- Seven studies of microcredit [Meager, 2020]
- The Oregon Medicaid experiment [Finkelstein et al., 2012]
- A study of cash transfers [Angelucci and De Giorgi, 2009]

Some analyses were robust, and others were not.

[1] https://github.com/rgiordan/zaminfluence

Can Dropping a Little Data Make a Big Difference?

We used our `R` package[1] examine a number of published analyses:

- Seven studies of microcredit [Meager, 2020]
- The Oregon Medicaid experiment [Finkelstein et al., 2012]
- A study of cash transfers [Angelucci and De Giorgi, 2009]

Some analyses were robust, and others were not.

What drives the variety of results?
We show that sensitivity to dropping small subsets is:

- Not (necessarily) caused by misspecification.
- Not (necessarily) caused by outliers.
- Not captured by standard errors.
- Not mitigated by large N.
- Primarily determined by the **signal to noise** ratio

  ... that is, the ratio of the measured effect size to data variability.

[1] `https://github.com/rgiordan/zaminfluence`

Links and references

Tamara Broderick, Ryan Giordano, Rachael Meager (alphabetical authors)
"An Automatic Finite-Sample Robustness Metric: Can Dropping a Little Data Change Conclusions?"
https://arxiv.org/abs/2011.14999

M. Angelucci and G. De Giorgi. Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review*, 99(1):486–508, 2009.

M. Angelucci, D. Karlan, and J. Zinman. Microcredit impacts: Evidence from a randomized microcredit program placement experiment by Compartamos Banco. *American Economic Journal: Applied Economics*, 7(1):151–82, 2015.

A. Finkelstein, S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. Newhouse, H. Allen, K. Baicker, and Oregon Health Study Group. The Oregon health insurance experiment: Evidence from the first year. *The Quarterly Journal of Economics*, 127(3):1057–1106, 2012.

R. Meager. Aggregating distributional treatment effects: A Bayesian hierarchical analysis of the microcredit literature. *LSE working paper*, 2020.